working paper
department
of economics

THE PERFORMANCE OF COAL-BURNING ELECTRIC GENERATING UNITS IN THE UNITED STATES: 1960-1980

Paul L. Joskow
Richard Schmalensee

Number 379 July 1985

massachusetts
institute of
technology

50 memorial drive
cambridge, mass. 02139

THE PERFORMANCE OF COAL-BURNING ELECTRIC GENERATING UNITS IN THE UNITED STATES: 1960-1980

Paul L. Joskow
Richard Schmalensee

Number 379 July 1985

The Performance of Coal-Burning Electric Generating Units In The United States: 1960-1980

Paul L. Joskow and Richard Schmalensee*

ABSTRACT

The factors that determine the thermal efficiency and reliability of coal-burning generating units are studied by applying recently developed techniques for dealing with panel data, allowing for the presence of unobservable unit-specific effects that may be correlated with observable variables, to a new and comprehensive data set. Existing econometric techniques are extended to allow for an unbalanced panel. Consistent and efficient estimates of the effects of unit age, vintage, scale, operating practices, and coal quality are obtained. Separate estimates are provided for two major technological groups. Some evidence is found that large utilities integrated into design and engineering obtain superior generating unit performance. The results have implications for the computation and evaluation of the life-cycle costs of generating electricity, the application of generating unit performance norms by regulators, the nature of technological change in steam-electric generating technology, and public policies toward mergers in the electric power industry.

July 1985

I. INTRODUCTION

This paper examines the major factors that influence the operating performance of coal-burning steam-electric generating units over time and space. We focus on two important aspects of generating unit performance: thermal efficiency and operating reliability. The analysis is motivated by several related issues and objectives:

a. Steam electric generating units are long-lived capital facilities. Economic decisions involving acquisition and replacement should depend in part on the expected performance of these facilities over long periods of time. While considerable research has focused on the construction costs of generating units and the costs of generating electricity at a point in time, there has been little systematic analysis of the actual life-cycle costs of these facilities or the important performance factors that influence these costs. Most economic analyses of the costs of generating electricity rely on engineering assumptions about generating unit performance, rather than on actual performance, or rely on observed performance only for the first few years of operation of a sample of generating units.[1]

b. Because poor generating unit performance means higher costs, regulators have become increasingly concerned with generating unit performance. Several regulatory agencies have introduced or are considering the introduction of performance "yardsticks" that can be used to evaluate the effectiveness with which individual utilities operate their generating units.[2] The basic idea

in many of these regulatory proposals is to compare actual performance of a particular regulated firm's units with some norm. The regulatory agency then uses the relationship between actual performance and the norm to determine allowable costs and rates for the regulated firm. Performance below the norm yields a financial penalty and performance exceeding the norm may, in some states, result in a financial reward. Developing an effective norm is necessarily difficult. It is rarely the case that one can find even one or two units operated by other utilities that are exactly like a unit subject to such an evaluation in all relevant dimensions. Not only is it difficult to match unit-specific characteristics, but many time-varying factors can be expected to affect observed performance at a point in time. Accordingly, a number of regulators are considering basing performance norms on statistical models that relate observed performance to a variety of time-invariant and time-varying unit characteristics.[3] It has not generally been recognized, however, that because such modeling generally involves the use of panel data with different numbers of observations on each unit, potentially complex specification and estimation problems must be solved to obtain consistent and efficient estimates of the parameters of the model.[4]

c. The basic steam cycle technology used to generate electricity has undergone a continuous evolution since the beginning of this century.[5] Technological innovation has made it possible for units with higher steam pressures and temperatures to be built. Higher steam temperature and pressure, at least theoretically, imply higher thermal efficiency.[6] Similarly, technological developments in both steam generation technology and transmission technology have made it feasible and potentially economically

desirable to build units of increasing size. And until the mid 1970's new units installed in any year had, on average, higher theoretical thermal efficiencies and were larger than those installed earlier.[7] Little systematic analysis has been done to examine whether the actual performance of these units is consistent with their theoretical thermodynamic properties. More important, the possibility that as engineers pushed out the technological frontier in the dimensions of steam temperature and pressure and of unit size, the reliability of units would fall, was generally not considered in performing evaluations of the economic desirability of continuing to build ever larger units designed to produce higher and higher pressure steam.[8] Since the mid-1970's, utilities appear to have retreated from the technological frontier in both the size and steam pressure dimensions.[9] Anectdotal evidence suggests that one reason for this change from historical trends has been the poor reliability of large units generally and the highest pressure (greatest theoretical thermal efficiency) units in particular.[10] Systematic statistical analysis of actual performance has been minimal, however.

d. There is wide diversity in the size of electric power companies in the U.S.[11] Unlike many other countries, electricity in the U.S. is not supplied by one or a handfull of large public or private enterprises. The typical electric utility in the U.S. must rely on third parties for design, construction and major engineering assistance with new generating units.[12] Such a utility will have a relatively small number of "similar" units operating and may not be able to take advantage of any economies of scale or experience in design, operation and maintenance that may be present. Since

we can observe the performance of the units of four large utilities with a large number of coal units and internal design and/or construction teams, we are in a position to test whether such economies may be present. In light of current regulatory policies that severly restrict mergers between electric utilities, the presence of such economies is an important public policy issue.[13]

We have put together a large panel data set on generating unit performance and the time-invariant and time-varying explantory variables that we hypothesize affect observed performance over time and space at the generating unit level. We are thus in a position to examine each of these issues in some detail. The nature of the problem and the data we rely on are particulary well suited to the application of recently developed econometric techniques for estimating models using panel data.[14] Because our data set is an unbalanced sample we develop and apply a relatively straightforward generalization of the techniques of Hausman and Taylor (1981) to the case of unbalanced panel data.

The paper proceeds as follows. The second section specifies the basic statistical model that we rely on and discusses the performance, time-invariant and time-varying variables of interest. The third section discusses the data employed in the study. The fourth section presents the econometric methods used. The fifth and sixth sections present the results. A summary of the results and their implications for the issues identified above concludes the paper.

II. THE MODEL

We are interested in examining the behavior of two performance variables. The first is a unit's thermodynamic efficiency. This is measured by the unit's gross heat rate (GHR); the number of btu's of fuel used to generate a Kwh of electricity.[15] The second is the unit's reliability. This is measured by the unit's equivalent availability factor (EAF); this is essentially the percentage of each year that a unit is available for operation at full capacity.[16] The higher is the thermal efficiency of a unit, *ceteris paribus*, the lower the cost of generating electricity. The greater the reliability of the unit, the more often the facility can produce output and the lower are the maintenance requirements, also reducing generating costs, *ceteris paribus*.

We want to estimate the effects of a number of unit-specific (time-invariant) and time-varying variables that we hypothesize affect observed performance. The time-invariant variables could include such things as unit vintage, unit size, construction cost, the specific technology embodied in the unit, an indicator for whether the unit is operated by one of four major utility companies that do their own engineering and construction work, etc. The time-varying variables that are hypothesized to affect observed performance include unit age, coal characteristics, maintenance activities and certain operating characteristics. We discuss the precise variables included in the analysis presently.

We hypothesize that the observed performance of a generating unit is a function of unit-specific characteristics that do not vary with time as well as operating characteristics that vary over time. Furthermore, some unit-

specific characteristics may be unobservable. Following Hausman and Taylor (1981), we assume that the determination of both GHR and EAF can be modeled as follows:

$$Y_{it} = X_{it}\beta + Z_i\gamma + \alpha_i + \eta_{it}, \qquad (1)$$

where β and γ are kx1 and gx1 vectors of coefficients associated with the observable time-varying (X_{it}) and time-invariant (Z_i) characteristics respectively. The disturbance η_{it} is assumed to be uncorrelated with the columns of (X,Z,α) and to have a zero mean and constant variance $\sigma^2(\eta)$ conditional on X_{it} and Z_i. The unobservable unit-specific effect α_i is assumed to be a time-invariant random variable, distributed independently across units, with variance $\sigma^2(\alpha)$.

If the α_i are uncorrelated with the columns of X and Z, one can obtain consistent estimates of β and γ using ordinary least squares (OLS) and consistent and efficient estimates using generalized least squares (GLS). But if the α_i are correlated with the columns of X and Z, OLS and GLS yield biased and inconsistent estimates of the parameters of interest. Fixed effects estimation still produces consistent estimates of β. But those estimates are inefficient, and this technique does not permit estimation of γ. For all the models examined in this study, the specification test presented by Hausman (1978, Sect. 3) decisively rejects the null hypotheses that α is uncorrelated with (X,Z). In order to obtain consistent and efficient estimates of both sets of parameters, we accordingly employ the techniques presented by Hausman and Taylor (1981), modified slightly to allow for the "unbalanced" nature of our data. These techniques, described more

fully in Section IV, permit one to treat some of the variables in (X,Z) as endogenous (i.e., correlated with α), to test the assumption that the remaining variables are exogenous (i.e., uncorrelated with α), and to obtain consistent and efficient (GLS/IV) estimates of β and γ even though some variables are endogenous.

We estimate separate versions of (1) with gross heat rate (GHR) and equivalent availability (EAF) as dependent variables (Y).[17] The units in our data base fall into two primary technological groups: subcritical units with steam pressures below about 2500 psi and supercritical units with steam pressures above 3206 psi. The latter represent the most recent development of the Rankine steam turbine technology. Most of the subscritical units in our sample have design steam pressures around 2400 psi; a smaller number have design pressures around 1800 psi. We estimate separate equations for subcritical and supercritical units because we felt a priori that they would exhibit different performance characteristics.[18] (We are able to reject the corresponding null hypotheses of coefficient identity; see Section IV.D) Given the relatively small number of subcritical units in the 1800 psi category, we estimate a single set of equations for all subcritical units and introduce a pressure dummy variable, PD/HIGH, equal to one for 2400 psi units and zero otherwise.

Given the thermodynamic implications of steam pressure, supercritical units should have heat rates about 2-3% lower than 2400 psi units, and 2400 psi units should have heat rates about 2-3% lower than 1800 psi units, all else equal.[19] The engineering literature provides no information that would allow us to make a priori predictions about differences in EAF across technologies. However, our discussions with utility engineers and the

interview results reported by Gordon (1983) suggest that as the industry pushed out the technological frontier, and especially as it moved to supercritical units, EAF's declined substantially.

The remainder of this Section defines the time-varying (X) and time-invariant (Z) characteristics that we expect to affect actual performance. All but two of these variables are used in both GHR and EAF equations. We also discuss the most likely sources of correlation between observable (X,Z) and unobservable unit-specific characteristics (α).

A. Time-Varying Characteristics (X).

AGE. We expect performance eventually to deteriorate as a unit ages. But units may go through a break-in period early in their lives, so that observed performance may actually improve during the first few years of operation. (The break-in period may be characterized by a high level of forced outages and derating or cycling of the facility, and we control for these factors separately -- see below). We have no a priori reason to choose a particular functional form for the aging profile of generating units, so we initially estimate the model allowing AGE (= calendar year minus year of initial operation) to enter with a high-order polynomial specification and report final results for the specification that exhausts the explanatory power of this variable.

COALBTU and COALSUL. The quality of coal burned will affect the operating performance of a generating unit. We have no way to identify all of the relevant coal characteristics that might be important.[20] We have data on two frequently-discussed characteristics: COALBTU = the btu content of the coal (measured as btu's per pound), and COALSUL = the sulfur content of the

coal (measured as a percentage). Other things equal, high btu fuel can be burned more efficiently than low btu fuel and should yield a lower heat rate. Higher btu coal also tends to be lower in ash content and other impurities that can foul boiler equipment, both reducing thermal efficiency and increasing the probability of outages and unit derating.[21]

The effects of COALSUL on operating performance are less clear a priori. On the one hand, sulfur is an impurity that should have a tendency to reduce operating performance, other things (including COALBTU) equal. On the other hand, during our sample period many units were forced to shift to the use of coal with a lower sulfur content to meet air pollution regulations. (Almost all of our units were subject to sulfur restrictions contained in State Implementation Plans, rather than the New Source Performance Standards). Regulatory constraints varied widely from unit to unit.[22] In the process of shifting to lower sulfur coal, utilities likely shifted to coal with several characteristics different from those contemplated in unit designs. This could lead to a deterioration in operating performance. If, other things equal, units observed to use coal with below average sulfur content tend to be those units that have been forced to shift the most from design coal specifications (which we cannot measure directly), higher observed sulfur content may be associated with better observed performance. In this case, COALSUL would be correlated with an unobservable unit-specific variable measuring deviations between design coal quality and actual coal quality. Correlations between COALBTU, COALSUL, and an unobservable variable could also arise because regional differences in the characteristics of indigenous coal are correlated with regional differences in design standards or operating practices.

EAF (GHR equations only). We expect performance to depend on how units are operated and maintained, but we cannot observe these practices directly. For the GHR equations, we have used the unit's contemporaneous EAF to measure the operating practices of interest. The lower was EAF during any period, the more time the unit was forced to operate at less than design capacity or not at all. Since heat energy must be expended to heat the boiler and other components when a unit is restarted, such deratings tend to produce lower measured heat rates. This is not really a causal argument; EAF is simply the best proxy we have for a unit's being operated (by choice or as a consequence of unplanned outages) in a way that reduces thermal efficiency. One might expect EAF to be correlated with a number of unobservable unit characteristics (α), including errors made at the design or construction stage.

OUTFAC (EAF equations only). In the EAF equations we introduce the unit's "output factor", OUTFAC, to measure the relevant operating practices. Output factor is defined as the actual kwh generation of the unit, expressed as a percentage of the maximum possible generation if the unit had been run at capacity whenever it was available. OUTFAC will be lower if a unit is cycled up and down to follow changes in load than if it is used as a base load plant and operated continuously at capacity. But cycling imposes more wear and tear on the equipment and makes it more likely that the unit will break down and be unavailable. Again, OUTFAC might be expected to be correlated with a variety of unobservable unit characteristics (α) that affect performance; "lemons" are more likely to be cycled than "stars".

B. Time-Invariant Variables (Z)

SCALE. Other things equal (in particular, steam temperature and pressure and fuel characteristics), the underlying thermodynamic properties of a Rankine steam cycle imply that increasing the size of the boiler should reduce the unit's heat rate, at least within some range.[23] The advantages of larger size should be more important at small scale than large scale. At very large scale heat rates may even begin to increase, particularly if very large units are to some extent experimental. In order to capture these effects, we estimate models allowing SCALE, measured as design capacity in megawatts (Mwe), to enter with a flexible polynomial specification. We generally get little if any increase in explanatory power with polynomials of order higher than two and accordingly report results for GHR with SCALE entered quadratically.

Engineering and economic studies of generating units traditionally assume that EAF is independent of unit capacity.[24] However, there is both "folk wisdom" and superficial empirical evidence drawn from average EAF's by unit size category that suggests that larger units have poorer availability than smaller units.[25] We also estimate the EAF equations allowing SCALE to enter with a flexible polynomial specification. A linear or quadratic specification generally exhausts the explanatory power of this variable.

VINTAGE. One would normally expect technological change to reduce GHR and to increase EAF over time. And, at least until the mid-1960's, a pattern of secular improvements in average thermal efficiency was observed.[26] However, by estimating separate equations for subcritical and supercritical units, and by using the dummy variable PD/HIGH to control for steam pressure differences among subcritical units, we control for the most important

"improvements" in technology during this period and examine changes over time in performance among units in the same "technological group." Secular improvements in thermal efficiency might still be observed in our sample, but during the period for which we observe performance (1969-1980), new plant designs had to be adapted to cope with increasingly stringent restrictions on sulfur, particulate, hot water and other emissions. These adaptations may have had the independent effect (given technological group) of raising GHR. Similarly, experience and improvements in technology should lead to increases in observed EAF, but design changes necessary to meet new environmental and safety regulations could lead to lower actual availability. Exactly how any secular improvements in "within group" technology balance out against deterioration in performance due to environmental regulation is an empirical question. As with SCALE, we estimate GHR and EAF equations allowing VINTAGE (= the year of initial operation minus 1959) to enter with a flexible polynomial specification, but find that either a linear or quadratic specification exhausts this variable's explanatory power.[27]

UD/AEP, UD/TVA, UD/SOCO, and UD/DUKE. We distinguish between two different types of utilities that own and operate generating units. The typical utility is relatively small, contracts infrequently with independent architect-engineers (AE) and constructors to design and build generating facilities and operates a relatively small number of units. A few large utilities both build numerous units and design and build these units using internal engineering and/or construction teams. Design and construction experience appears to lead to lower initial construction costs.[28] We are interested in testing whether large experienced utilities with internal engineering staffs also achieve better operating performance than does the

typical utility. We have identified four large coal-burning utilities that do their own engineering and design work and frequently do their own construction as well: American Electric Power (AEP), Southern Company (SOCO), Tennessee Valley Authority (TVA) and Duke Power (DUKE). Four utility dummy (UD) variables are employed that equal one if the unit was built by the corresponding large utility and equal zero otherwise. If there is any advantage to experience and internal control, as is sometimes suggested in the literature, these utilities should exhibit superior performance in one or both of the dimensions we analyze.

Construction Cost. It is natural to consider including the initial construction cost of a unit in these equations as well. Simple static theory would suggest that, all else equal, if a utility spends more money it will get a unit that performs better. However, in previous work using a sample of subcritical units built during the 1960's, we were unable to find a quantitatively or statistically significant tradeoff between a unit's intrinsic performance attributes and its initial construction cost. (Schmalensee and Joskow (1985)) Other work with the present data set suggests that cost relationships among units with different steam pressures (and design thermal efficiencies) is quite complex. (Joskow and Rose (1985)) Furthermore, utility engineers with whom we have spoken have suggested that conscious tradeoffs between initial construction costs and performance are rarely made within technological groups. Nevertheless, since we had construction cost data for most of the units in this sample, we tried several specifications in which construction cost per kw of capacity was a unit-specific variable. In all cases but one, unit construction cost had no

explanatory power; in the remaining case its sign was incorrect. In light of this, we dropped construction cost from our analysis.

C. Unobservable Characteristics (α)

As noted above, unobservable unit-specific characteristics that affect performance may well be correlated with the coal quality variables, COALBTU and COALSUL, and with the proxy measures of operating practices, EAF and OUTFAC. VINTAGE may also be correlated with α, especially for units embodying the newest (supercritical) technology, since design changes are likely to have occurred over time in response to both refinements in technology and changing environmental and safety regulations. The estimation procedure that we employ allows us to test for "endogenety" associated with left-out unobservable variables and to obtain consistent estimates where this is a problem.

III. THE DATA SET

We began construction of our data set with a comprehensive list of coal-fired generating units with capacities of at least 100 Mwe that began commercial operation between 1960 and 1980. (See Joskow and Rose (1985) for details.) These units accounted for about 95% of all coal-fired generating capacity installed during this period. For these units we have obtained data on SIZE, VINTAGE, steam characteristics (which divided units between the subcritical and supercritical samples and provided PD/HIGH for units in the latter), and architect-engineer (AE) (which provided values for the four UD variables). We merged this data set with information collected by the National Electric Reliability Council (NERC) covering the period

1969-1980, from which we took annual observations by generating unit on EAF, OUTFAC, and capacity factor. Capacity factor is defined as the ratio of actual generation to the product of unit capacity and the number of hours in the year ("period hours"); it was used (along with capacity and period hours) to compute actual generation. The NERC data did not cover about 25% of the units in the Joskow/Rose data base. Because data are missing for certain years, and because some units began operating during the sample period, the number of observations varies from unit to unit.

This merged data set was then merged again with FPC/FERC data on fuel utilization by generating unit, derived from responses to FPC/FERC Forms 67 and 423. The FPC/FERC data base included information on the quantity of fuel burned by each unit, COALBTU, and COALSUL. Using COALBTU and the quantity of coal burned, along with the generation figures derived from the NERC data, we calculated GHR for each unit/year observation. This merging process reduced the size of the sample further, both because of differences in units covered and because many obvious errors in the FPC/FERC data set made it necessary to drop additional observations.

Table 1 gives the means and standard deviations for each basic variable in the two GHR and EAF samples, as well as the number of units and unit/year observations in each. Note that the supercritical units in our data set tend to be newer and larger than the subcritical units, to be slightly more efficient, and to have distinctly lower availability. The differences in numbers of observations between the GHR and EAF samples arises because missing observations or obvious reporting errors occur most frequently in the reports on fuel use by generating unit, which are used to calculate GHR.

Table 1. - Means and (Standard Deviations) of Basic Variables Employed

Variable	Subcritical Units GHR Sample	Subcritical Units EAF Sample	Supercritical Units GHR Sample	Supercritical Units EAF Sample
Dependent (Y)				
GHR	9436. (564.)		9239. (464.)	
EAF		75.32 (16.1)		66.66 (15.7)
Time-Varying (X)				
AGE	8.927 (5.10)	9.026 (5.13)	5.956 (3.97)	5.989 (3.98)
COALBTU	11081. (1283.)	11032. (1284.)	11614. (799.)	11590. (904.)
COALSUL	2.037 (1.10)	1.945 (1.10)	2.239 (.916)	2.154 (.959)
EAF	75.39 (15.7)		67.36 (15.4)	
OUTFAC		78.06 (11.6)		82.26 (9.57)
Time-Invariant (Z)				
SCALE	349.7 (179.)	342.2 (174.)	699.7 (230.)	698.9 (224.)
VINTAGE	7.268 (4.90)	7.208 (5.00)	10.60 (3.64)	10.52 (3.59)
PD/HIGH	.7301 (.444)	.6945 (.461)		
UD/AEP	.0162 (.126)	.0135 (.116)	.1728 (.378)	.1597 (.367)
UD/TVA	.0316 (.175)	.0271 (.162)	.0517 (.222)	.0474 (..213)
UD/SOCO	.0829 (.276)	.0748 (.263)	.1004 (.301)	.0934 (.291)
UD/DUKE	.0337 (.181)	.0283 (.166)	.0502 (.219)	.0460 (.210)
Number of Units	181	225	82	89
Total Observations	1423	1699	677	739

Note: Figures in parentheses are standard deviations.

IV. ECONOMETRIC METHODS

This Section describes the econometric methods used to obtain consistent and efficient estimates of various versions of equation (1). These methods are relatively straightforward generalizations of the techniques of Hausman and Taylor (1981), referred to as HT in what follows, to the case of unbalanced panel data. Accordingly, we follow their presentation and notation and omit details of proofs.

A. GLS Estimation

Suppose the sample contains data on N units, with T_i observations on unit i, and let S be the sum of the T_i (= NT in the balanced case). Suppose that observations in (1) are ordered first by unit and then by time, so that α and the columns of Z are Sx1 vectors having N blocks, each with T_i identical entries, for i = 1,...,N. Let P and D be SxS block-diagonal matrices with N blocks. The i^{th} block of P is a $T_i x T_i$ matrix, all elements of which equal $1/T_i$, and the i^{th} block of D is T_i times the $T_i x T_i$ identity matrix. P is idempotent of rank N, $Q \equiv I - P$ is idempotent of rank S-N, and QP=PQ=0. With data grouped by units, multiplication by P transforms a vector of observations into a vector of unit-specific means, and multiplication by Q produces a vector of deviations from unit-specific means. (P and Q generalize the matrices P_V and Q_V, respectively, in HT.)

With this notation, the disturbance covariance matrix of (1) can be written as

$$\Omega = \sigma^2(\eta)I + \sigma^2(\alpha)DP. \tag{2}$$

Let Θ be an SxS diagonal matrix, in which the T_i diagonal elements corresponding to observations on unit i are all equal to

$$\theta_i = \{\sigma^2(\eta)/[\sigma^2(\eta)+T_i\sigma^2(\alpha)]\}^{1/2}. \tag{3}$$

One can then show that multiplication of (1) by

$$\Omega^{-1/2} = \Theta P + Q = I - (I-\Theta)P \tag{4}$$

yields an equation with scalar disturbance covariance matrix $\sigma^2(\eta)I$. The transformed equation can thus be consistently and efficiently estimated by OLS as long as α and η are independent of (X,Z). Multiplication by $\Omega^{-1/2}$ simply multiplies the observations on unit i in Z by θ_i (since Z = PZ) and subtracts $(1-\theta_i)$ times the i^{th} unit-specific mean from the corresponding observations in X. The GLS estimates β_{GLS} and γ_{GLS} are thus easily computed if consistent estimates of the disturbance variances in (3) are available.

In order to obtain the necessary variance estimates, one employs within-unit (fixed effects) and between-unit regressions as in HT. First, multiplication of (1) by Q yields the within-unit equation, which relates deviations of Y and X from the corresponding unit-specific means (since QZ = 0). Its disturbance covariance matrix is $\sigma^2(\eta)Q$. Application of OLS to this relation yields a consistent estimate of β, β_W, and division of the resultant sum of squared residuals by (S-N) yields a consistent estimate of $\sigma^2(\eta)$. Second, multiplication of (1) by P yields the between-unit equation, which relates the unit-specific means of Y to those of the variables in X and to Z.

The between-unit disturbance covariance matrix is $[\sigma^2(\eta)+\sigma^2(\alpha)D]P$. (Note that this transformed equation has T_i identical observations on the unit-specific means corresponding to unit i, for i=1,...,N.) Application of least squares yields another consistent estimate of β, β_B, and division of the corresponding sum of squared residuals by N yields a consistent estimate of $[\sigma^2(\eta)+(S/N)\sigma^2(\alpha)]$. Substituting the estimate of $\sigma^2(\eta)$ derived from the within-units regression, a consistent estimate of $\sigma^2(\alpha)$ is obtained.

B. Basic Specification Test

A key maintained hypothesis in GLS estimation is that $E(\alpha|X,Z)=0$. If this hypothesis is correct, β_W and β_B are consistent, but β_{GLS} is more efficient than either, while if this hypothesis is incorrect, only β_W is consistent. HT present three large-sample χ^2 tests of the null hypothesis $E(\alpha|X,Z)=0$ involving differences between pairs of these estimates and prove that they are numerically identical in the balanced case.[29]

These tests are also numerically identical in the unbalanced case as well, but one is the clear choice on computational grounds. As in the balanced case, the OLS covariance matrix from the within-units regression is not a consistent estimate of $V(\beta_W)$, the covariance matrix of β_W. This is easily corrected by a degrees of freedom adjustment: OLS divides the sum of squared residuals by (S-k) in estimating the disturbance variance but, as noted above, consistency requires division by (S-N) (or (S-N-k)). (At least among our students, negative "χ^2" statistics are often produced by failure to make this correction. Intuitively, the correction is necessary because within-unit regressions are equivalent to fixed-effects models with N unit-specific dummy variables.) In balanced samples, the OLS covariance matrix from the between-units regression is a consistent estimate of $V(\beta_B)$, but this

is not true in the unbalanced case. Computation of such an estimate in the unbalanced case is fairly involved.[30]

Accordingly, the χ^2 test using β_{GLS} and β_W and the corresponding covariance matrices is much the simplest of the three HT tests to employ. (Note that the same estimate of $\sigma^2(\eta)$ should be used to compute both $V(\beta_W)$ and $V(\beta_{GLS})$ for numerical consistency.) Following Hausman (1978, sect. 3), this test can be performed most easily as a χ^2 test of the null hypothesis $\delta=0$ in the following regression:

$$\Omega^{-1/2}Y_{it} = (\Omega^{-1/2}X_{it})\beta + (\Omega^{-1/2}Z_i)\gamma + (QX_{it})\delta + \varepsilon_{it}. \tag{5}$$

That is, one simply adds the deviations of the X's from their unit-specific means to the transformed (for GLS estimation) version of equation (1). In interpreting the results of this test, it is important to bear in mind that the independence of η and the columns of (X,Z,α), which is necessary for consistency, is part of the maintained hypothesis.

C. GLS/IV Estimation and Testing

If the basic specification test implies $E(\alpha|X,Z) \neq 0$, consistent estimation is still possible if correlation with α occurs only in a sufficiently small subset of the variables X and Z. Specifically, suppose $X = (X_1|X_2)$, where X_1 is Sxk_1, X_2 is Sxk_2, and the columns of X_1 are asymptotically uncorrelated with α. Similarly, let $Z = (Z_1|Z_2)$, where X_1 is Sxg_1, Z_2 is Sxg_2, and the columns of Z_1 are asymptotically uncorrelated with α. Then, as long as the order condition for identification, $k_1 \geq g_2$, is satisfied (along with the corresponding rank condition), the elements of β

and γ can be consistently estimated. Moreover, if $k_1 > g_2$, the null hypothesis $E(\alpha|X_1,Z_1) = 0$ can be tested.

If $k_2 > 0$ or $g_2 > 0$, the between-unit regression does not yield a consistent estimate of $\sigma^2(\alpha)$. The within-unit regression is used as in GLS to obtain β_W, which is consistent, and a consistent estimate of $\sigma^2(\eta)$. Let d be the Sx1 vector of unit-specific means of the residuals from this regression, stacked as usual:

$$d = P[Y - X\beta_W]. \tag{6}$$

If $g_2 = 0$, least-squares estimation of $d = Z\gamma$ yields a consistent estimate of γ, γ_W. If $g_2 > 0$ <u>and</u> $k_1 > g_2$, two-stage least-squares (TSLS) applied to this equation, with X_1 and Z_1 as instruments, produces such an estimate. Given γ_W, one can compute the Sx1 residual vector

$$e = Y - X\beta_W - Z\gamma_W. \tag{7}$$

Then $(e'e)/S$ is a consistent estimator of $[\sigma^2(\eta)+\sigma^2(\alpha)]$, and the required estimate of $\sigma^2(\alpha)$ follows immediately.[31]

Given a consistent estimate of $\Omega^{-1/2}$, HT show that transformation of (1) and application of TSLS yields consistent and efficient estimates of β and γ. No new instrumental variables are needed as long as $k_1 \geq g_2$. HT note that this technique works because only the time-invariant component of the disturbance (α) is correlated with (X_2,Z_2). This permits the variables in X_1 to do double duty: since $X_1 = PX_1 + QX_1$, and the two components are orthogonal, PX_1 can be used as an instrument for Z_2, while QX_1 serves as an instrument for X_1. Because of the structure of the model, it is simplest

(particularly with large unbalanced samples) to compute TSLS estimates in the classical two-step fashion, following Appendix B in HT.

The first step in these computations is to obtain "fitted values", $\hat{X}_2$ and $\hat{Z}_2$ corresponding to the endogeneous variables, X_2 and Z_2, respectively. Let $(\widehat{PX_2})$ be the fitted values from regressions of the columns of (PX_2) on (PX_1) and Z_1. Then $\hat{X}_2$ is given by

$$\hat{X}_2 = (\widehat{PX_2}) + QX_2. \tag{8}$$

(Note that QX_2 cannot be correlated with α.) Similarly, $\hat{Z}_2$ is obtained as the fitted values from regressions of the columns of Z_2 on (PX_1) and Z_1. The second step begins with substitution of $\hat{X}_2$ for X_2 and $\hat{Z}_2$ for Z_2 in (1). (One can show that the rank condition $k_1 \geq g_2$ is necessary for this substitution to yield a data matrix of full column rank, just as in more conventional applications of TSLS.) Then, exactly as in GLS, the resulting equation is transformed by pre-multiplication by $\Omega^{-1/2}$, and OLS is employed to compute estimates of β and γ. It is important to note that the estimates of the disturbance variance and (thus) the coefficient covariance matrix produced by OLS in the second step are inconsistent; as in any appplication of TSLS, one must use actual rather than fitted values of X_2 and Z_2 to obtain consistent estimates.

If $k_1 > g_2$, the χ^2 test presented in HT's Proposition 3.4 can used be to test the maintained hypothesis $E(\alpha|X_1,Z_1) = 0$. If $(S-k) > (k_1-g_1)$, one can follow Hausman (1978, Sect. 3) and perform this test in a regression framework. But the natural procedure, which involves adding (QX_1) to the (second-stage) regression equation used to compute the GLS/IV coefficient

estimates and applying OLS, will not work here. It is easy to show that the columns of $\hat{Z}_2$ (as defined above) are linear combinations of the columns of X_1, Z_1, and (QX_1). In order to avoid using generalized inverse routines (which most regression packages lack), the unrestricted equation to be compared with the original second-stage equation must be formed by adding (QX_1) and <u>deleting</u> $\hat{Z}_2$. The χ^2 statistic comparing these two regressions then has (k_1-g_2) degrees of freedom, exactly as in HT's Proposition 3.4. (Note that a consistent estimate of $\sigma^2(\eta)$ must be used in computing this test statistic.)

D. Testing Coefficient Stability

For each of the models discussed in Section V, using both GLS and GLS/IV estimation, we test the null hypothesis that subcritical and supercritical units have identical parameters. (When subcritical and supercritical specifications differ, we employ the minimal specification that includes both as special cases.) All of these χ^2 (large-sample Chow) tests reject the null hypothesis at conventional significance levels; the uninteresting details are omitted to save space. The relevant submatrices of the Q matrix estimated in order to apply GLS to the pooled sample must be used in estimating subsample relations for testing purposes. Further, it follows from the analysis of Lo and Newey (1983) that in GLS/IV estimation one must compute separate first-stage fitted values for each of the two subsamples and simply stack these to obtain the fitted values for pooled estimation. Finally, the relevant sums of squared residuals for computing the χ^2 test statistic are those from the second-stage regressions computed using the fitted values, $\hat{X}_2$ and $\hat{Z}_2$. (As above, however, a consistent

estimate of $\sigma^2(\eta)$, which these sums of squared residuals do not yield, must be used in computing the χ^2 statistic.)

V. ECONOMETRIC RESULTS: GROSS HEAT RATE

Table 2 contains the results for the heat rate (GHR) equations estimated for subcritical and supercritical units. We report estimates produced by OLS, GLS, and GLS/IV, along with the relevant χ^2 statistics for the specification tests performed on the GLS and GLS/IV estimates. Both GLS equations fail the basic specification test. Happily, GLS/IV estimates of equations with COALSUL and EAF, the variables a priori most likely to be correlated with α, treated as endogenous, yield specification test statistics that point toward acceptance of the null hypothesis.

A. Subcritical Units

The coefficient estimates for subcritical units are broadly consistent with our expectations. A quartic in AGE fits the data quite well, and the coefficients are not particularly sensitive to estimation method. The GLS and GLS/IV estimates suggest that heat rate improves for roughly four years after initial operation and then deteriorates for the rest of a unit's life. Thus, the data indicate a quantitatively important break-in period with regard to thermodynamic efficiency; GHR comes to exceed its value when AGE=0 only when AGE=9. On average, a 20-year-old unit's heat rate has risen by about 940 btu/Kwh from its lowest value; this is about 10% of the sample mean. This is quite significant, since, as we noted above, the *ceteris paribus* difference in theoretical thermodynamic efficiencies between 1800 psi

Table 2. - Estimates of Gross Heat Rate (GHR) Equations

	Subcritical Units			Supercritical Units		
	OLS	GLS	GLS/IV	OLS	GLS	GLS/IV
AGE	-170.0	-140.4	-139.9	41.64	44.75	41.09
	(4.14)	(4.53)	(4.50)	(9.04)	(11.6)	(10.2)
$(AGE)^2$	34.00	28.88	28.79			
	(3.93)	(4.46)	(4.43)			
$(AGE)^3$	-2.247	-1.826	-1.811			
	(3.33)	(3.62)	(3.59)			
$(AGE)^4$	.0526	.0412	.0407			
	(3.04)	(3.18)	(3.14)			
COALBTU	-.0632	-.0530	-.0519	-.1037	-.1041	-.0998
	(5.57)	(2.93)	(2.69)	(5.28)	(4.13)	(3.59)
COALSUL	-7.612	-5.372	-10.40*	12.27	90.98	16.80*
	(0.60)	(0.30)	(0.42)	(0.73)	(3.69)	(0.52)
EAF	-8.635	-3.981	-2.760*	-7.380	-4.169	-3.275*
	(9.75)	(5.15)	(3.40)	(8.14)	(5.11)	(3.85)
Constant	10861.	10200.	10079.	10118.	9094.	10258.
	(56.5)	(39.6)	(36.7)	(35.9)	(18.1)	(24.9)
SCALE	-1.053	-.7737	-.6449	-.7314	-.9506	-.9652
	(2.62)	(1.01)	(0.81)	(2.17)	(1.59)	(1.34)
$(SCALE)^2$	13.16[a]	8.915[a]	8.206[a]	6.316[a]	6.938[a]	6.604[a]
	(2.96)	(1.16)	(0.98)	(3.03)	(1.94)	(1.53)
VINTAGE	45.85	55.18	55.40	143.9	204.8	117.5
	(8.91)	(7.84)	(7.22)	(9.37)	(5.86)	(3.93)
$(VINTAGE)^2$				-5.324	-6.361	-4.932
				(7.50)	(5.37)	(3.76)
PD/HIGH	-299.6	-217.4	-217.8			
	(6.32)	(3.10)	(2.84)			
UD/AEP	-364.6	-419.8	-440.6	-278.3	-317.5	-276.7
	(3.57)	(1.88)	(1.78)	(6.38)	(4.10)	(3.14)
UD/TVA	-174.9	-91.75	-77.41	-416.0	-392.1	-287.9
	(2.11)	(0.52)	(0.40)	(5.06)	(2.72)	(1.69)
UD/SOCO	-10.27	-41.08	-45.58	170.0	132.7	152.1
	(0.22)	(0.48)	(0.48)	(3.62)	(1.73)	(1.63)
UD/DUKE	-441.7	-501.4	-521.2	-700.7	-669.6	-655.7
	(6.01)	(3.10)	(2.91)	(10.7)	(6.10)	(4.71)
$\sigma(\alpha)$	-	302.7	334.3	-	180.1	231.8
Std. Error	470.8	348.0	678.5	336.9	269.5	346.2
Spec. Test	-	$\chi^2(7)=28.6$	$\chi^2(5)=2.8$	-	$\chi^2(4)=37.2$	$\chi^2(2)=2.3$

Notes: Figures in parentheses are absolute values of t-statistics. (Since they do not take into account variance components, the OLS t-statistics are inconsistent. The GLS and GLS/IV t-statistics are computed using the consistent estimates of $\sigma(\eta)$ from within-units regressions: 340.7 for subcritical units, and 263.3 for supercritical units.) Starred variables are treated as endogenous in GLS/IV estimation.

subcritical units and 3500 psi supercritical units, which span the range of technological advance since 1960, is only around 6%.

We also find that there has been a significant secular deterioration in the performance of units as they entered service over time. Newer units built more recently are less efficient than units that entered service twenty years ago, other things equal (including unit age). Allowing the year of initial operation to enter with a higher order polynomial did not yield a significant reduction in unexplained variation or significant coefficients for the higher order terms. Since subcritical technology was reasonably mature at the beginning of our sample period, this suggests that design changes, perhaps in response to environmental restrictions, have led to lower performance over time. The GLS and GLS/IV estimates indicate that the VINTAGE-related difference in GHR between the oldest and the newest units in the sample is about 11.7% of the sample mean of GHR.

The OLS estimates indicate that SCALE also affects thermodynamic efficiency, as the engineering literature suggests, but these effects are insignificant in the GLS and GLS/IV equations. All three estimates of SCALE coefficients indicate that heat rate is minimized at about 400 Mwe (which is also about the mean size of subscritical units installed between 1960 and 1980), but the variation in heat rates from smallest to largest units in the sample is only about 1.4% of the sample mean of GHR according to the GLS and GLS/IV coefficients.[32]

As we expected, units that burn coal with a higher heat content have lower heat rates. According to the GLS and GLS/IV estimates, from lowest to highest value of COALBTU, the range in expected heat rate is about 3.6% of the sample mean of GHR. The estimated coefficients of COALSUL are negative

and never significant. Without unobservable environmental restrictions correlated with sulfur content, we would have expected a positive coefficient. The negative sign and lack of significance may reflect approximate cancellation of the two factors discussed in Section II.

The EAF variable, which we use as a proxy for the operating characteristics of a unit, is negative and highly significant even when it is treated as endogenous (correlated with α). Units that are derated a lot, and thus operate relatively often at less than optimum design capacity or are out of service entirely, exhibit poorer heat rates than units that are not subject to substantial forced outages and derating. This effect is not large, however: the GLS/IV estimate indicates than an increase in EAF of two sample standard deviations would lower GHR by about 0.9% of its sample mean. Note also that treating EAF as endogenous lowers its coefficient by about 30%, while allowance for variance components produces a 54% drop.

Units rated at 2400 psi (PD/HIGH = 1) have significantly lower heat rates than units rated 1800 psi, as the basic thermodynamic properties of a Rankine steam cycle would predict. The (GLS and GLS/IV) difference of 217 btu/Kwh, about 2.3% of the sample mean of GHR, is roughly what would be predicted from steam tables.

Finally, the four utilities that do their own design and engineering work and have a relatively large number of coal-fired units uniformly exhibit lower heat rates, other things equal. The difference is substantial only for AEP and DUKE, however.

B. Supercritical Units

The coefficient estimates for supercritical units follow a pattern that is qualitatively similar to that observed for subscritical units. The

performance of supercritical units also deteriorates as they age. Higher-order polynomial terms in AGE added nothing to this model; we find no evidence of a break-in period. Thermodynamic performance seems to begin deteriorating almost immediately, and a unit's heat rate is estimated (GLS/IV) to rise on average by just under 9% of the sample mean of GHR during 20 years of service.

We found clear evidence of a non-linear effect of VINTAGE. Unlike the subscritical units, for a considerable period of time (from 1960 through 1971 according to the GLS/IV estimates) new units exhibited lower heat rates than older units, all else equal. But since the mid-1970's at the latest, a secular deterioration in the performance of the newest units is evident. (The OLS estimates put the turning point at 1973; the GLS coefficients make it 1975.) Since supercritical technology was relatively new at the beginning of our sample period, the initial improvements in performance probably reflect significant technological progress that dominated the forces that led to lower performance for subcritical units. By the early 1970's, however, this progress apparently slowed or ceased, and thermal efficiency of new units began to decline, perhaps as a result of efforts to accomodate new environmental regulations. VINTAGE is estimated to have a substantial effect on the performance of supercritical units: the GLS/IV estimates indicate a decrease in GHR by about 6% of the sample mean from 1960 to 1971, followed by an 8% increase between 1971 and 1980.

The estimated effects of coal characteristics and unit size are similar to our estimates for the subcritical units. Increases in COALBTU are again found to improve efficiency, as expected. COALSUL is again insignificant, though its coefficient is now positive. A comparison of GLS/IV estimates

indicates that the GHR of supercritical units is roughly twice as sensitive to changes in COALBTU is the GHR of subcritical units. As for subcritical units, the SCALE coefficients become insignificant as we move from OLS to GLS/IV. The GLS/IV estimates indicate that thermal efficiency is maximixed at a capacity of 730 Mwe, just above the sample mean of SCALE, with the maximum SCALE-related differences in GHR within the sample amounting to only about 1% of the sample mean.

The coefficient of EAF is again negative and significant, and it declines substantially in absolute value when its possible endogeneity is allowed for. Supercritical units seem slightly more sensitive to variations in EAF than subcritical units. All else equal, units owned by AEP and DUKE have significantly lower heat rates than other units, and the coefficient of UD/TVA is negative and significant at 5% on a one-tailed test in the GLS/IV estimates. The Southern Company again fails to exhibit above-average performance.

Our analysis of the actual thermodynamic efficiency of subcritical technology indicated that units designed to operate at higher steam pressures (2400 psi) are more efficient than units designed to operate at lower steam pressures (1800 psi). The magnitude of the difference in observed performance, other things equal, is approximately equal to the theoretical difference drawn from engineering calculations (2 to 3%). The motivation for moving to higher pressure supercritical technology was to increase thermodynamic efficiency (i.e. reduce the heat rate) further. Our results suggest that the actual performance of these units falls short of these design engineering expectations. If we evaluate the two (GLS/IV) heat rate equations at the means of the independent variables for the subcritical sample we find that the predicted heat rate for supercritical units is

actually higher than for subcritical units. If we evaluate the equations at the means of the independent variables for the supercritical sample we find that the predicted values for subcritical and supercritical units are about the same. Even in the latter case, 2400 psi subcritical units have lower predicted heat rates than supercritical units. It is only for the post-1970 vintages of supercritical units that we find the expected lower predicted heat rates from this technology, reflecting the improvement in supercritical technology over time.

VI. ECONOMETRIC RESULTS: EQUIVALENT AVAILABILITY (EAF)

Table 3 presents the results of OLS, GLS, and GLS/IV estimates of the parameters of equations determining equivalent availability. The basic specification test clearly rejects the null hypothesis $E(\alpha|X,Z) = 0$ for both GLS equations. Here the variables most likely correlated with α are OUTFAC and the coal characteristics. For the subcritical sample, treating OUTFAC and COALBTU as endogenous seems to solve the problem. (In contrast to the GHR equations, treating COALSUL as endogenous in this sample has essentially no effect on the χ^2 statistics.) The situation is more complex for the supercritical sample. Treating OUTFAC and COALBTU as endogenous gives a χ^2 specification test statistic of 9.0 with two degrees of freedom. Adding VINTAGE to the set of endogenous variables reduces the χ^2 to 4.3 with one degree of freedom. Adding COALSUL instead gives a $\chi^2(1)$ statistic of 3.3. This last test does not reject the null hypothesis at the 5% level, though it does reject at the 10% level. Since a model with OUTFAC, COALBTU, and COALSUL treated as endogenous almost passes the specification test, and VINTAGE is both economically and statistically a good candidate for

Table 3. - Estimates of Equivalent Availability (EAF) Equations

	Subcritical Units			Supercritical Units		
	OLS	GLS	GLS/IV	OLS	GLS	GLS/IV
AGE	-.3159	-.3243	-.1647	-.4458	-.4153	-.1866
	(1.05)	(1.20)	(0.60)	(2.49)	(2.45)	(0.96)
$(AGE)^2$	-.0291	-.0277	-.0351			
	(2.05)	(2.18)	(2.72)			
COALBTU	-1.806[a]	2.988[a]	45.64[a]*	15.48[a]	17.04[a]	2.037[a]*
	(0.57)	(0.63)	(4.48)	(2.93)	(2.12)	(0.11)
COALSUL	-.8510	-.7292	-1.302	-.9434	-.3594	2.824*
	(2.47)	(1.51)	(2.37)	(1.58)	(0.48)	(1.84)
OUTFAC	.2855	.2852	.2853*	.5876	.4973	.4182*
	(9.27)	(8.55)	(7.46)	(9.82)	(8.06)	(5.86)
Constant	80.75	73.81	26.71	8.559	14.40	30.26
	(1.60)	(11.3)	(2.26)	(0.82)	(1.15)	(1.24)
SCALE	-.0346	-.0318	-.0252	-.0355	-.0376	-.0423
	(11.2)	(6.24)	(3.99)	(2.76)	(2.14)	(1.71)
$(SCALE)^2$				.1284[a]	.1638[a]	.1887[a]
				(1.60)	(1.51)	(1.25)
VINTAGE	-1.619	-1.500	-1.477	1.162	.9911	1.166*
	(5.35)	(3.42)	(2.83)	(4.55)	(3.25)	(2.61)
$(VINTAGE)^2$	.0811	.0757	.0953			
	(5.36)	(3.64)	(3.84)			
PD/HIGH	-1.719	-1.931	-6.531			
	(1.82)	(1.22)	(3.07)			
UD/AEP	5.017	5.366	3.049	9.069	9.035	7.561
	(1.66)	(0.96)	(0.44)	(5.31)	(3.84)	(2.08)
UD/TVA	-1.124	-1.948	.2648	1.526	-2.174	-6.708
	(0.46)	(0.44)	(0.05)	(0.46)	(0.47)	(1.00)
UD/SOCO	2.240	1.997	-1.764	1.596	1.763	3.966
	(1.63)	(0.92)	(0.65)	(0.85)	(0.70)	(1.12)
UD/DUKE	6.775	6.791	5.323	11.50	11.20	15.28
	(3.14)	(1.69)	(1.06)	(4.40)	(3.08)	(2.89)
$\sigma(\alpha)$	-	6.88	8.89	-	4.98	8.17
Std. Error	14.09	12.32	19.59	14.01	12.97	14.52
Spec. Test	-	$\chi^2(5)=26.6$	$\chi^2(3)=3.6$	-	$\chi^2(4)=22.3$	-[b]

Notes: Figures in parentheses are absolute values of t-statistics. (Since they do not take into account variance components, the OLS t-statistics are inconsistent. The GLS and GLS/IV t-statistics are computed using the consistent estimates of $\sigma(\eta)$ from within-units regressions: 12.17 for subcritical units, and 12.73 for supercritical units.) Starred variables are treated as endogenous in GLS/IV estimation.

[a]Coefficients of COALBTU and $(SCALE)^2$ have been multiplied by 10^4 for presentation purposes.

endogeneity, we feel confident in assuming that treating all four of these variables as endogenous yields consistent estimates. Unfortunately, it also yields a just-identified model ($k_1 = g_2 = 1$), so that this assumption cannot be tested.

A. Subcritical Units

All three sets of estimates show a large, monotonic, accelerating decline in availability with unit AGE; we find no evidence of a break-in period. Over the first 20 years of a unit's life, the GLS/IV estimates predict a decline in EAF of 18 percentage points, about 24% of the sample mean. The estimated VINTAGE effect is highly significant and rather surprising. Beginning with units coming on line in 1960, we observe a ceteris paribus decline in EAF for new units enterring commercial operation in each year until 1967 (GLS/IV) or 1969 (OLS and GLS). Thereafter, later vintages show higher EAF's, until by 1980 the VINTAGE effect is above its 1960 value. The peak-to-trough difference in EAF is about 17 percentage points according to the GLS/IV estimates. There is no compelling explanation for this result, but we offer two hypotheses. First, Joskow and Rose (1985) found that the real construction cost of new coal-burning generating units declined until the later 1960's and then increased thereafter, other things equal. It is possible that the secular reduction in costs during the 1960's led to lower unit reliability and that the subsequent secular increases in costs are a result of efforts to increase reliability. Second, the secular deterioration in thermal efficiency that we find during the 1970's might be a consequence of design changes made to newer units in an effort to improve the poor reliability of earlier units. As we indicated above, however, we have tried to account for variations in constructon costs and interactions between thermal efficiency and reliability in this and

related work (see above and Schmalensee and Joskow (1985)) and have been unable to find firm statistical support for these hypotheses. Since the interelationship between construction costs, unit performance and operating practices may be more complex than what we have allowed for to date in our modeling efforts, we consider these to be reasonable hypotheses that are worth further exploration.

The OLS results suggest that the BTU content of coal burned is not an important determinant of availability, and the estimated coefficient is negative. We expected just the opposite, since higher BTU coal tends to be lower in ash and other impurities that can lead to more frequent boiler maintenance and failures in the boiler system. Both GLS and GLS/IV coefficients of COALBTU are positive, and the latter is significant. The perverse OLS result thsu appears to be a consequence of the failure of OLS to appropriately account for variance components and endogeneity problems. The GLS/IV estimate implies that an increase in COALBTU of two sample standard deviations will raise EAF by about seven percentage points. COALSUL has the expected negative coefficient, which is significant in both OLS and GLS/IV estimates. Sulphur seems a less important determinant of availability than COALBTU; the GLS/IV coefficient indicates that an increase in COALSUL of two sample standard deviations will lower EAF by only about two percentage points.

Our estimates also suggest quite strongly that larger subcritical units are subject to a significantly higher probability of outage and derating than smaller units. This is consistent with the non-econometric evidence discussed in Section II, above. According to the GLS/IV estimates, the EAF for an 800 Mwe unit is about 13 percentage points lower than for a 300 Mwe

unit, all else equal. This is a very large and economically significant difference in the ratio of actual, effective capacity to nominal capacity.

The operating characteristics of the unit, for which OUTFAC serves as a proxy, also appear to be important. A unit that operates continuously when it is available (OUTFAC=100) has an EAF 14 percentage points above that of a unit that has an output factor of only 50%. The coefficient of OF is estimated quite precisely and is insensitive to choice of estimation method. Cycling units up and down appears to lead to significant wear and tear and ultimately to equipment failures.

It is also interesting to note that 2400 psi units (PD/HIGH=1) appear to have lower availabilities than 1800 psi units. The GLS/IV estimate of this difference is 6.5 percentage points, 8.7% of the sample mean of EAF and almost four times the OLS estimate. This difference in EAF's is substantial both absolutely and relative to the 2.3% differences in GHR's discussed in Section V. It also turned out that, other things equal, supercritical units have EAF's that are about 12% lower than subcritical units (see below). Our findings are thus fully consistent with the notion, discussed in Section II, that units close to the scale/pressure/temperature frontier of technology are less reliable than those with less adventurous designs.

Finally, there is some evidence to support the hypothesis that the EAF's for the four large utilities identified above are higher than the EAF's for a typical utility for subcritical units. Three of the four coefficients of the UD variables are positive and those for AEP and DUKE are relatively large numerically, but none of the coefficients is estimated very precisely when GLS/IV is applied.

B. Supercritical Units

The results for supercritical units are broadly similar to those we obtain for the subscritical units, but, as with the GHR equations, there are some interesting differences. We find units age approximately linearly, with no evidence of a break-in period. Suprisingly, the AGE coefficient is insignificant in the GLS/IV equation, probably because of the assumed endogeneity of VINTAGE and the built-in negative correlation between AGE and VINTAGE in our data. The OLS and GLS estimates imply that supercritical units experience only about an 8.5 percentage-point decrease in EAF over 20 years, about half the estimated deterioration in subcritical availability.

Newer units appear to have higher availabilities than older units, all else equal. Thus, as we found in the heat rate equations for supercritical units, design changes in newer units appear to have led to better performance as experience was gained with the technology. The estimated VINTAGE effect is quite large: about 12 percentage points after 10 years according to GLS/IV. This suggests that the early supercritical units were experimental to an important extent and as a consequence were real lemons in terms of reliability and availability.

As with the subcritical units, availability seems to deteriorate as supercritical units get larger, though the GLS/IV coefficients of SCALE are estimated rather imprecisely. Despite the quadratic term, all three estimates imply that increases in SCALE lower EAF within the sample range, except possibly (according to GLS and GLS/IV) for the very largest units observed. The GLS/IV estimates imply that the largest unit in the sample (SCALE=1300) has an EAF about 11 percentage points lower than the smallest unit (SCALE=325) as a consequence of scale differences alone.

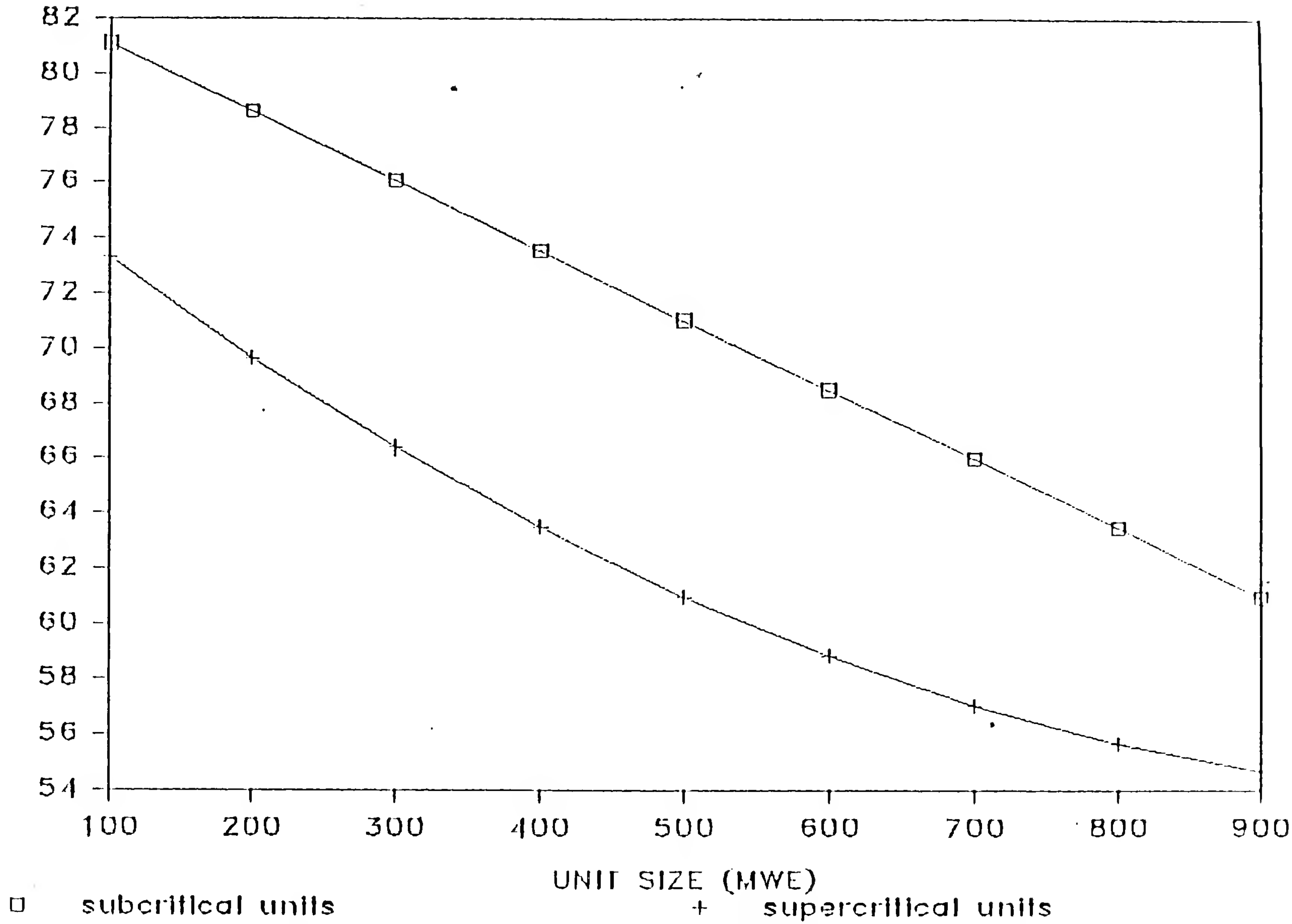
Increases in the BTU content of coal burned are estimated to enhance unit availability, as expected, though this effect is insignificant in the GLS/IV estimates. The coefficient of COALSUL is never significant, and its sign is at odds with expectations based on engineering considerations in the GLS/IV estimates. But, as we noted in Section II, environmental restrictions could give rise to a positive correlation (with no causal significance) between COALSUL and performance.

Even when its possible correlation with unobservable unit-specific effects is allowed for, the output factor appears to be a statistically and quantitatively significant determinant of unit availability. Supercritical units seem to be more sensitive to cycling than subcritical units; an increase in OUTFAC from 50 to 100 is associated with an increase in EAF of 21 percentage points for supercitical units, as compared to 14 percentage points for subcritical units (GLS/IV estimates). There is also evidence that both AEP and DUKE achieve superior availability of their supercritical units.

Finally, when we compare the predicted EAF for supercritical units with that for subcritical units we find that supercritical units exhibit significantly lower levels of reliability. When we evaluate the two EAF equations (GLS/IV) at the means of the independent variables of the subcritical sample, we find that supercritical units have an EAF about 10 percentage points (14%) lower than supercritical units (See Figure 1). Evaluating the equations at the means of the independent variables for the supercritical sample yields a predicted EAF for supercritical units that is 7 percentage points (10%) lower than that predicted for subcritical units. The EPRI Technical Assessment Guide assumes that 500 Mwe subcritical and supercritical units will achieve EAF's of about 74%. This is very close to observed performance for subcritical units, but far off for supercritical

FIGURE #1

PREDICTED EQUIVALENT AVAILABILITY BY UNIT SIZE
AND TECHNOLOGY



units (64% estimated vs. 74% assumed). Furthermore, EPRI assumes that there is a reduction in EAF of about 3 percentage points as we move from 500 Mwe to 1000 Mwe units. The actual falloff is closer to 10 percentage points. Overall, it appears that supercritical units have performed far worse in terms of reliability than engineering analyses have assumed.

VII: SUMMARY AND CONCLUSIONS

It is useful to discuss these results in light of the issues that motivated this analysis. It is quite clear that the performance of steam electric generating units varies widely, but systematically, over time and space. Appropriate economic calculations of electricity costs and economic evaluations of the desirability of units with different steam conditions and different sizes are likely to be sensitive to these performance characteristics and should be incorporated in such analyses. Performance observed for the first few years of a unit's life are not good indicators of life-cycle performance. Assumptions that availability is independent of unit size and technical characteristics are inconsistent with observed performance. Failure to account for these variations in performance is likely to lead to incorrect economic calculations.

Unit performance in both the heat rate and availability dimensions deteriorates significantly as units age. While larger units tend to have slightly lower heat rates than smaller units, they also exhibit much poorer reliability. Larger sizes for generating units must be justified by construction cost savings rather than operating cost savings. And there is substantial evidence that larger units <u>are</u> less costly to build than smaller units (Joskow and Rose). But because larger units have much poorer

availabilities than smaller units, the apparent overall economic advantage of larger units due to lower construction costs may disappear when the costs of poor reliability are factored in.

From the perspective of regulators interested in developing norms for evaluating the performance of the utilities under their jurisdiction, there are a number of implications of the results presented here. Sensible performance criteria must be sensitive to the unit-specific and time-varying characteristics of the facilities for which norms are being established. Simple averages or even simple grouping procedures are not likely to yield meaningful norms. Unit age, vintage, technological characteristics and coal characteristics all affect observed performance in important ways and should be controlled for. While statistical analyses such as those performed here should be useful for establishing such norms, regulators should also recognize that the estimates obtained are, at least for some variables, quite sensitive to the estimating technique employed. Efforts to apply models such as this to develop norms must take careful account of the econometric issues that we have discussed and apply appropriate econometric techniques to deal with them.

The nature of technological change that has characterized steam generation technology over the past ten to twenty years appears to be quite complex. Increases in the steam pressures of generating units have led to improvements in thermal efficiency. But at least with regard to the movement to supercritical technology, these improvements did not occur very quickly. Early vintages of supercritical technology were, on average, no more efficient, and apparently somewhat less efficient, than contemporary state-of-the-art subcritical units. Furthermore, these improvements have been achieved at what may be a substantial cost. Other things equal, as steam

pressure has been increased, unit availability has decreased. Those units with the highest theoretical thermal efficiency have the poorest reliability. Finally, to the extent that poor reliability is a good proxy for the tendency of units to be operating at other than optimum levels from a thermal efficiency perspective, at least some of the theoretical thermal efficiency gains will be eroded. It is evident that utilities have had to pay a substantial reliability penalty for efforts that have been made to extend technological capabilities. In the last several years it appears that utilities have retreated from both the maximum size and steam pressure frontier (Gordon (1983), Joskow and Rose (1985)). One likely reason for this is the high cost of poor reliability.

We find the vintage effects that we have estimated to be particularly puzzling. At least during the 1970's there is substantial evidence that "within-technology" thermal efficiency declined. This is the case for subcritical units during the entire sample period. Thus there appears to be a sort of negative technological change in the thermal efficiency dimension. It is possible that the deterioration in thermal efficiency is a consequence of design changes necessitated by new environmental regulations during the 1970's. But we offer this only as a hypothesis that must be subjected to further investigation. The evidence on reliability, at least during the 1970's, suggests that newer units within each group have achieved better reliability than older units, something that we might expect as a result of increasing experience and the higher cost of more recent units. It is possible that the deterioration in thermal efficiency associated with newer units during this period of time may not be a consequence of environmental regulations, but rather may reflect design changes aimed at improving reliability. These changes may have necessitated reducing the thermal

efficiency of the units. Again, we offer this as a hypothesis that should be subjected to further analysis.

Finally, the results obtained here suggest that the four large utilities with internal engineering and design teams generally achieve better performance than does the typical utility. While the importance and statistical significance of this size and experience effect varies between technologies and between the two performance attributes, taken as a whole, the results suggest that organizational considerations have an impact on the performance of these facilities. These results should at least raise some questions about the wisdom of public policies that have restricted mergers between utilities and discouraged the formation of more entities such as these. We have presented elsewhere additional evidence that suggests that this policy has probably been costly (Joskow and Schmalensee (1983), Joskow and Rose (1985)).

FOOTNOTES

*Massachusetts Institute of Technology.
The authors are indebted to the U.S. Department of Energy, the MIT Center for Energy Policy Research, and the Department of Economics energy research fund for financial support, to John Landon for providing data to us, and to Lant Pritchett for superb research assistance. We have benefited greatly from discussions with Jerry Hausman, Jim Powell, and Lew Perl. The usual waiver of liability applies.

1. Perl's (1982) unpublished study is an exception

2 . See Johnson (1985), page 41, and National Regulatory Research Institute (1981).

3. See Johnson (1985), p.61.

4. Corio (1982), Perl (1982) and Landon and Huettner (1984) report the results of econometric analyses of generating unit performance, but rely on OLS estimation and, in the case of Corio, on a much smaller data base.

5. See Bushe (1981), Cootner and Lof (1965) and Joskow and Rose (1985).

6. See especially Cootner and Lof (1965), Chapters 3, 5 and Appendix A.

7. See Joskow and Rose (1985), Tables 1 and 2.

8. Ling (1964), Cootner and Lof (1965), Bushe (1981), Wills (1978), and Cowing (1974) appear to take no account of variations over time, scale and technology in unit reliability. The EPRI Technical Assessment Guide (1982, B-55) assumes that subcritical and supercritical units of equivalent size have the same reliability. 1000 Mwe supercritical units are assumed to have reliability levels less than 5% lower than 500 Mwe units.

9. Joskow and Rose (1985).

10. Gordon (1983).

11. See Joskow and Schmalensee (1983), Chapter 2.

12. See Joskow and Rose (1985).

13. See Joskow and Schmalensee, Chapters 2, 7 and 14.

14. See Hausman and Taylor (1981).

15. At the generating <u>plant</u> level, most discussions of heat rates refer to the <u>net</u> heat rate. The net heat rate is based on the quantity of electricity sent from the plant to the grid. The gross heat rate is based on the amount of electricity generated by the turbines. The difference is accounted for by electricity used within the plant itself to run equipment and to provide lighting. There is no way to calculate net heat rates at the generating <u>unit</u> level. For our purposes, in any case, the gross heat rate is what is

relevant since it gives us a pure measure of the performance of the boiler and generating equipment.

16. Adjustment is made for "partial outages," which reduce effective capacity. See National Electric Reliability Council (n.d.) for the formal definition of EAF.

17. In principle, of course, one could enhance estimation efficiency by allowing for cross-equation correlations between the α_i and, possibly, the η_{it}. In order to do this properly, however, it would not only be necessary to extend the Hausman-Taylor GLS/IV techniques (as generalized in Section IV, below) to the multiple equations case, but the nature of our data set would require a further extension to the case of differing numbers of units and unit/year observations among equations. Neither of these would be a simple task, and we feel justified in considering them beyond the bounds of the present study. (Consider the treatment of the second issue in the much simpler fixed-effects context in the Appendix to Schmalensee and Joskow (1985).)

18. Supercritical units (generally 3500 psi) are designed to have lower heat rates than subcritical (since 1960 either 2400 or 1800 psi), other things equal. The EPRI Technical Assessment guide assumes that supercritical units will have heat rates 2.6% lower than subcritical units at average load levels. (EPRI Technical Assessment Guide at B-55). Gordon (1983) suggests that supercritical units have experienced unusual availability problems.

19. See Cootner and Lof (1965) and EPRI Technical Assessment Guide (1982, B-55).

20. We can observe BTU content, ash content, moisture content and sulfur content. A large fraction of the variation in BTU content is explained by variations in ash and moisture content. We cannot observe coal grindability, coal size or the chemical content of the coal. Some coal contracts specify more than a dozen coal quality attributes. See Joskow (1985).

21. The EPRI Technical Assessment Guide (1982) assumes that units using low BTU lignite have heat rates about 4% higher than units using High BTU bituminous coal at average load levels (page B-51). Interviews with utility power plant engineers indicated that plant availability is sensitive to variations in coal quality. See Joskow (1985).

22. See for example Gollop and Roberts (1983).

23. See <u>Mark's Standard Handbook for Mechanical Engineers</u> (8th Edition), pages 9-54 to 9-56.

24. See footnote 8, <u>supra</u>.

25. See, for instance, Gordon (1983), Loose and Flaim (1980), and Joskow and Schmalensee (1983, p. 48).

26. See Edison Electric Institute, Statistics of the Electric utility Industry/1982, p. 32 and Historical Statistics of the Electic Utility Industry, page 115.

27. Because environmental regulations changed over time, one might expect the performance of existing units to change for reasons unrelated to unit design or aging. But, since observation date, AGE, and VINTAGE are linearly dependent, such an effect cannot be separately identified; if present, it is reflected in coefficients of AGE and VINTAGE terms.

28. Joskow and Rose (1985) find evidence of experience effects in unit construction for both utilities and architect-engineers.

29. For the sake of completeness, it should be noted that the HT proof of the equivalence of the three possible χ^2 tests (their Proposition 2.2) is not invalidated by what appears to be an error in the unnumbered equation at the top of p. 1381 to which they refer in the course of that proof. (Compare Madalla (1971, p. 343) and consider the dimensionality of V_B and V_W.) Only the definition of Δ in the key identity in the last line on p. 1382 is affected, and that does not affect the proof.

30. Let $W = (X|Z)$. Then the covariance matrix of the between-units coefficient vector, $[(\beta_B)',(\gamma_B)']'$, is given by

$$V_B = \sigma^2(\eta)(W'W)^{-1} + \sigma^2(\alpha)(W'W)^{-1}(W'DW)(W'W)^{-1}.$$

In the balanced case, D is a scalar matrix, and the last term simplifies so as to imply the consistency of the usual OLS estimator of V_B.

31. Following HT (p. 1384) exactly, one would use (Pe)'(Pe)/S as an estimator of $[(N/S)\sigma^2(\eta)+\sigma^2(\alpha)]$. This is numerically equivalent to the approach in the text but involves somewhat more computation, particularly in the unbalanced case.

32. This is consistent with the theoretical thermodynamic relationship between heat rate and unit size. See Mark's Standard Handbook for Mechanical Engineers, page 9-55.

REFERENCES

Bushe, Dennis, "An Empirical Analysis of Production and Technology Using Heterogeneous Capital: Thermal Electric Power Generation," unpublished PhD dissertation, New York University, 1981 (University Microfilms, Ann Arbor, Michigan).

"Coal-Fired Plants: Efficient and Reliable," EPRI Journal (December 1979), 18-24.

Cootner, Paul H. and George O.G. Lof, Water Demand for Steam Electric Generation, Resources For The Future, Washington, DC, 1965.

Corio, Marie, R., "Heat Rates and Availability: The Use of Econometric Analysis in Comparing Electric Generating Unit Performance," mimeographed, National Economic Research Associates, Inc., Summer 1981.

————————, "Why is the Performance of Electric Generating Units Declining?" Public Utilities Fortnightly, April 29, 1982, 3-8.

Cowing, Thomas, "Technical Change and Scale Economies in an Engineering Production Function: The Case of Steam Electric Power," Journal of Industrial Economics, December, 1974.

Electric Power Research Institute, Technical Assessment Guide, May 1982.

Gollop, Frank and Mark Roberts, "Environmental Regulations and Productivity Growth: The Case of Fossil-Fueled Electric Power Generation," Journal of Political Economy, August, 1983.

Gordon, Robert J., "The Productivity Slowdown in the Steam-Electric Generating Industry," mimeographed, Northwestern University, February 1983.

Hausman, Jerry A., "Specification Tests in Econometrics," Econometrica 46 (November 1978), 1251-1272.

Hausman, Jerry A., and Taylor, William E., "Panel Data and Unobservable Individual Effects," Econometrica 49 (November 1981), 1377-1398.

Johnson, Leland, "Incentives to Improve Electric Utility Performance," Rand Corporation, March 1985.

Joskow, Paul L., "Vertical Integreation and Long Term Contracts: The Case of Coal-Burning Electric Generating Units," Journal of Law, Economics and Organization (in press).

Joskow, Paul L., and Nancy L. Rose, "The Effects of Technological Change, Experience and Environmental Regulation on the Construction Cost of Coal-Burning Generating Units," Rand Journal of Economics, Spring 1985.

Joskow, Paul L. and Richard Schmalensee, Markets for Power (Cambridge: MIT MIT Press, 1983).

Landon, John and David Huetner, "Utility Performance Evaluation," A Report to the Rate Research Committee of the Edison Electric Institute, September 18, 1984.

Ling, Suilin, Economies of Scale in the Steam Electric Power Generating Industry. (Amsterdam: North-Holland, 1964).

Lo, Andrew, and Whitney Newey, "A Residuals-Based Wald Test for the Single Linear Simultaneous Equation," mimeographed, Harvard University, September 1983.

Loose, Verne W., and Teresa Flaim, "Economies of Scale and Reliability: The Economics of Large versus Small Generating Units," Energy Systems and Policy 4 (1980), 37-56.

Madalla, G.S., "The Use of Variance Components Models in Pooling Cross Section and Time Series Data," Econometrica 39 (1971), 341-358.

National Electric Reliability Council, Generating Availability Data System, Ten Year Review, 1971-1980, Report on Equipment Availability. (Princeton: National Electric Reliability Council, n.d.).

Perl, Louis J., "The Current Economics of Electric Generation from Coal in the U.S. and Western Europe," mimeographed, National Economic Research Associates, Inc., October 1982.

Schmalensee, Richard, and Paul L. Joskow, "Estimated Parameters as Independent Variables: An Application to the Costs of Electric Generating Units," Working Paper 1575-84, Sloan School of Management, MIT, February 1985.

Wills, Hugh R., "Estimation of a Vintage Capital Model for Electricty Generating," Review of Economic Studies, October, 1978.

2-5-90

Date Due